FoodChangeLens: CNN-based Food Transformation on HoloLens

Shu Naritomi
Department of Informatics
The University of Electro-Communications
Tokyo, Japan
n1610480@edu.cc.uec.ac.jp

Ryosuke Tanno
Technology Development
NTT Communications Corporation
Tokyo, Japan
r.tanno@ntt.com

Takumi Ege Keiji Yanai
Department of Informatics
The University of Electro-Communications
Tokyo, Japan
{ege-t,yanai}@mm.inf.uec.ac.jp

***Abstract*—In this demonstration, we implemented food category transformation in mixed reality using both image generation and HoloLens. Our system overlays transformed food images to food objects in the AR space, so that it is possible to convert in consideration of real shape. This system has the potential to make meals more enjoyable. In this work, we use the Conditional CycleGAN trained with a large-scale food image data collected from the Twitter Stream for food category transformation which can transform among ten kinds of foods mutually keeping the shape of a given food. We show the virtual meal experience which is food category transformation among ten kinds of typical Japanese foods: ramen noodle, curry rice, fried rice, beef rice bowl, chilled noodle, spaghetti with meat source, white rice, eel bowl, and fried noodle. Note that additional results including demo videos can be see at https://negi111111.github.io/FoodChangeLensProjectHP/**

***Index Terms*—Deep Learning, Convolutional Neural Network, Generative Adversarial Networks, HoloLens, Food Image Transfer**

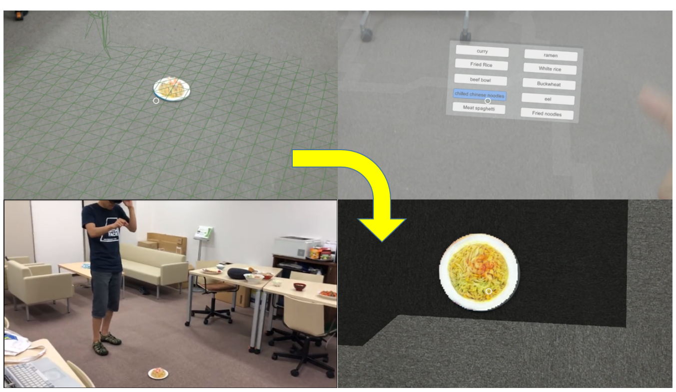

Fig. 1. Food Category Transformation in AR space using HoloLens (Fried rice to Chilled Chinese noodle).

I. Introduction

We develop an application, called “FoodChangeLens”, for transforming food category in the AR space using Microsoft mixed reality glasses Microsoft HoloLens. The aim is to transform food category in the real space to food category selected by a user in the AR space and provide an virtual meal experience that makes meals more enjoyable.

In this work, we combined image generation with mixed reality, so that it is possible to transform food category considering three-dimensional geometry. For transforming food category, we use Conditional CycleGAN which extended CycleGAN [5] and for considering three-dimensional geometry, we use HoloLens which enable to capture mesh data of real space and pastes a texture by taking correspondence of each vertex.

Our demonstration begins with a single subject wearing HoloLens as shown in the lower left part of Figure 1 and performs according to the following flow. Firstly, in the scanning phase, mesh data is given by scanning the surroundings. Secondly, the food category is selected from the category selection window by air tap gesture. Finally, a gazed food object are transformed to selected food category by air tap gesture.

The main contributions of this demo paper is summarized as follows:

1) We combine an image transformation and mixed reality using HoloLens.
2) We show the food category transformation which is completely a new application in the AR space.

II. Related Work

Zhu et al. proposed a method to train an image transformation network using unpaired training samples which consists of two domains of image samples such as color images and corresponding grey-scale images [5]. They introduced a cycle consistency loss for training, and successfully trained an image transformation network which transform an original-domain image to the other-domain image keeping rough shape structure unchanged.

In this work, we use a cycle consistency loss to train a model. In a food domain, with this loss function, we can transfer a food image to the other category of a food image keeping the original food image structure unchanged.

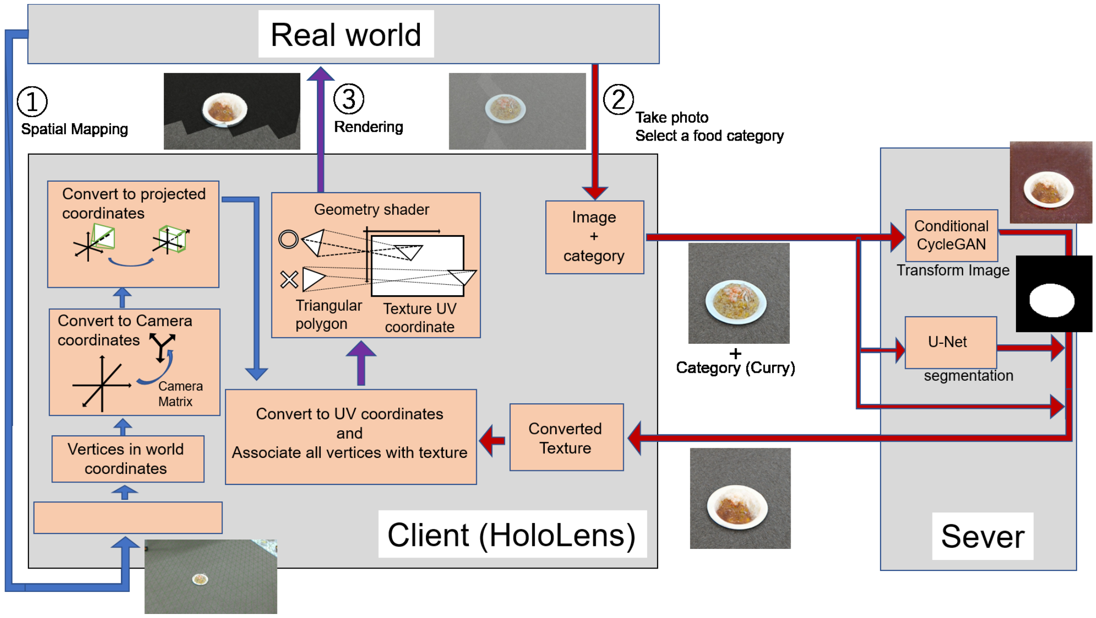

Fig. 2. The overview of our entire system of food category transformation using HoloLens.

III. IMPLEMENTATION

In this work, we overlay transformed images to food objects on the three-dimensional space, by mapping transformed images to objects in the AR space, which can convert the appearance of a real food rather than a food photo.

A. Overview of the Proposed System

A schematic diagram of our entire system is shown in Figure 2, and the flow of processing is shown below.

1) Mesh data of real space is given by using HoloLens that can capture mesh data of three-dimensional space by scanning real space.
2) The image taken by the HoloLens built-in camera is sent to the server and the transformed image is received from the server.
3) Using the transformation matrix and the projection matrix of the world coordinates obtained from HoloLens, all vertices of all the mesh data, at the time the picture was taken, are converted from the world coordinates to the camera coordinates, and it is converted to the projected coordinates. The coordinates converted to the projected coordinates are normalized and corresponds to the coordinates of the image received from the server. Then, it is confirmed whether each vertex has a corresponding point or not, and information for passing to the GPU is created as an array.
4) The data saved in (3) is handed over to the shader, a triangle polygon is received in the geometry shader, and it is confirmed whether all three vertices have corresponding points to the texture. If one of the three vertices does not correspond to the texture, it returns a transparent value. If there is correspondence for all the vertexes, the transformed image is taken as a texture.

B. Food Category Transfer

We show the network of Conditional CycleGAN (cCycleGAN) in Figure 3 which is a conditioned extension of CycleGAN. cCycleGAN can convert a given image to the image which belongs to the indicated category by adding a conditional input to an image transformation network of CycleGAN [5]. To use a conditional vector effectively, in cCycleGAN we added Auxiliary Classifier Loss to the discriminator [3]. The discriminator of cCycleGAN classifies not only real or fake but also category of images, which is the same approach to [1].

Also, in order to transform only food area, the food area segmentation was performed by using U-Net [4]. We use the UEC-FOOD100 dataset [2] as a training data for U-Net that can identify food area and non-food area. Since the UEC-FOOD100 dataset has a bounding-box annotation, a pseudo food area mask is generated using GrabCut, and it is used for training the network as ground-truth.

Finally, using both the transformed image by cCycleGAN and the food segmentation result by U-Net, we transformed only food area by applying the transformed result only to the area estimated as food.

IV. DEMONSTRATION

Our demonstration begins with a single subject wearing Microsoft HoloLens as shown in the lower left part of Figure 1 and performs according to the following flow. Firstly, in the scanning phase, mesh data is given by scanning the surroundings. Secondly, the food category is selected from the category selection window by air tap gesture. Finally, a gazed

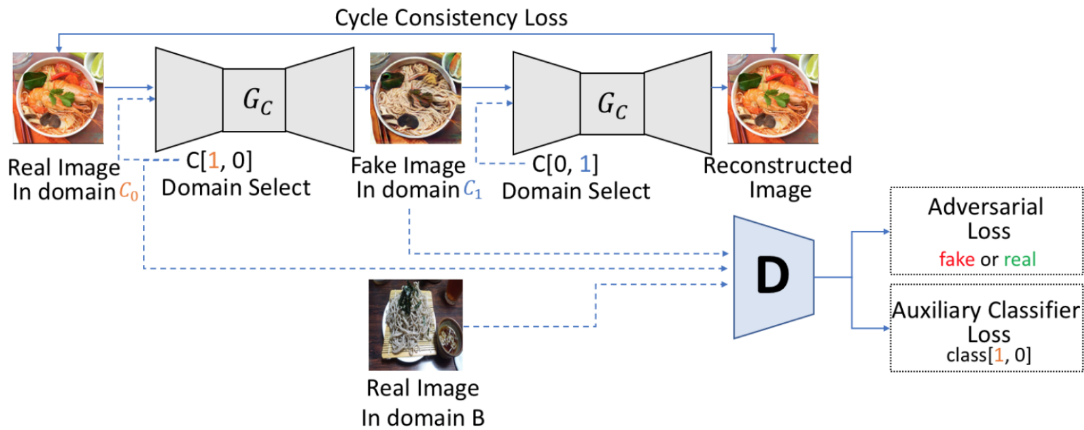

Fig. 3. The architecture of the Conditional Cycle-GAN.

food object is transformed to selected food category by air tap gesture.

V. CONCLUSIONS

In this work, we combined image generation with mixed reality, so that it is possible to transform food category considering three-dimensional geometry. For transforming food category, we use Conditional CycleGAN which extended CycleGAN and for considering three-dimensional geometry, we use HoloLens. This has potential applications of virtual meal experience which makes meals more enjoyable.

As a future work, we plan to implement advanced visualization and real time transformation.

Note that additional results including demo videos can be see at https://negi111111.github.io/FoodChangeLensProjectHP/

Acknowledgments: This work was supported by JSPS KAKENHI Grant Number 15H05915, 17H01745, 17H05972, 17H06026 and 17H06100.

REFERENCES

[1] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proc. of IEEE Computer Vision and Pattern Recognition (CVPR)*, 2018.

[2] Y. Matsuda, H. Hoashi, and K. Yanai. Recognition of multiple-food images by detecting candidate regions. In *Proc. of IEEE International Conference on Multimedia and Expo (ICME)*, 2012.

[3] A. Odena, C. Olah, and J. Shlens. Conditional image synthesis with auxiliary classifier gans. In *Proc. of the International Conference on Machine Learning (ICML)*, 2017.

[4] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proc. of arXiv:1505.04597v1*, 2015.

[5] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2017.